

Transactional Data vs. Hypermedia for Data in Motion

Transactional Data



VS.

Hypermedia



Transactional Data vs. Hypermedia for Data in Motion

- Transactional Data (Structured Data)
- Hypermedia
- Unstructured Data
- Unstructured Data Formats and File Types
- The Challenges of Unstructured Data
- Web Services, ReST
- Data in Motion
- Metadata and Tagged Content
- Role of the Data Architect



Transactional Data vs. Hypermedia for Data in Motion

Transactional Data

- In a SOA context, **a transaction is the movement or exchange of data that completes a single interaction.**
- While a singleton SOA transaction does not necessarily have to be fully ACID, for most use cases an ACID transaction is highly desirable.
- ACID:
 - ✓ **Atomic** – a transaction is treated as a complete unit, e.g. treated as “all or nothing”
 - ✓ **Consistent** – a transaction and its data remain in a consistent state, before and after
 - ✓ **Isolated** – a single transaction and its data remain unaffected by other consumers and services
 - ✓ **Durable** - if a transaction fails, it can be recovered and the data returns to its original state, e.g. backed out, compensation
- One exception to the ACID convention for a singleton transaction are when that transaction participates as part of a larger, decomposed or federated unit of work (e.g. a Federated Transaction, often enabled by a SOA Service Orchestration).
- **Transactional Data can be thought of as one or more data elements and values that represent, define or participate within a single unit of work.**



Transactional Data (Structured Data)



Transactional Data vs. Hypermedia for Data in Motion

Transactional Data

- For SOA, a SOAP Web Service is perhaps the most common and easy to describe type of transaction.
- The SOAP message is encoded using XML and as you can see from the example, individual and discretely defined data elements are included in the SOAP “body”.

```
<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
  xmlns:SOAP-ENC="http://schemas.xmlsoap.org/soap/encoding/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <SOAP-ENV:Body>
    <m:Request xmlns:m="http://Widget-Example.com/Inventories"
      <m:ItemID>5678</m:ItemID>
      <m:Desc>Blue Stick Pen</m:Desc>
      <m:Version>2</m:Version>
      <m:Length>8.0 Inch</m:Length>
      <m:Color>Blue</m:Color>
      <m:Diameter>0.25 Inch</m:Diameter>
      <m:Weight>1.5 oz</m:Weight>
    </m:Request>
  </SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```

```
<m:ItemID>5678</m:ItemID>
<m:Desc>Blue Stick Pen</m:Desc>
<m:Version>2</m:Version>
<m:Length>8.0 Inch</m:Length>
<m:Color>Blue</m:Color>
<m:Diameter>0.25 Inch</m:Diameter>
<m:Weight>1.5 oz</m:Weight>
```



Transactional Data vs. Hypermedia for Data in Motion

Transactional Data

- In short, Transactional Data is very much about discretely defined data elements, data values and metadata.
- If we look at the metadata that describes transactional data, we can easily see that it represents “structure”.
 - ❑ Element Names
 - ❑ Data Types
 - ❑ Lengths
 - ❑ Decimalization (Scale and Precision)
 - ❑ Type Patterns (REGEX or other)
 - ❑ Modality
 - ❑ Relationships with other Elements
 - ❑ Valid Values
 - ❑ And so on ...
- That is, “structure” as a well-defined, repeatable and consistent, form and format.



Hypermedia



Transactional Data vs. Hypermedia for Data in Motion

Hypermedia

- Hypermedia extends the notion of transactional data to one that includes and combines things like:

Images



Video



Audio



- Hypermedia is more about a set of information that represents an “experience”.
- As you browse today’s Web, it is quite common to find and experience Hypermedia as embedded or linked and downloadable images, audio files and movies.



Unstructured Data

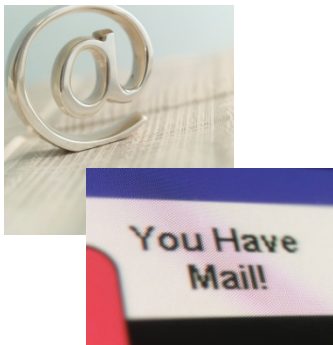


Transactional Data vs. Hypermedia for Data in Motion

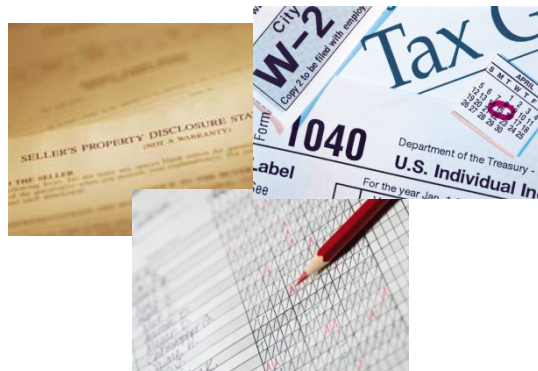
Unstructured Data

- While the Web presents a number of very common examples of hypermedia such as images, audio and video, these are only a subset of unstructured data.
- Unstructured data is a superset that further extends the notion of hypermedia.
- In the typical enterprise we also have other forms of unstructured data such as:

Email



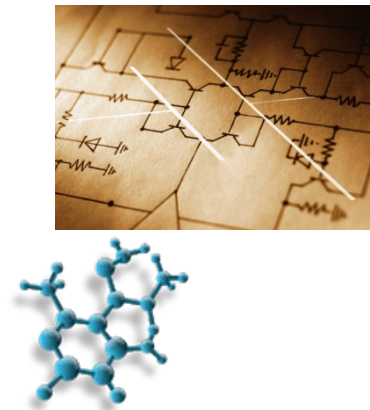
Documents



Charts and Graphs



Diagrams and Schematics



Transactional Data vs. Hypermedia for Data in Motion

The Challenges of Unstructured Data

- Unstructured data tends to be a coarse-grained, large set of encoded data (“bits and bytes”) that are treated as a unit.
- The analogy is often one of a “file” that is downloaded or in some manner rendered and decoded for consumption.
- In a SOA context, unstructured data can be embedded within a transaction or it can be linked and referenced by a transaction.



Unstructured Data Formats and File Types



Transactional Data vs. Hypermedia for Data in Motion

Unstructured Data Formats and File Types

A few common examples of file types and extensions for unstructured data.

Type	Name	File Extension
Images	Bitmap Image File	.bmp
Images	Graphical Interchange Format File	.gif
Images	JPEG Image File	.jpg
Images	Portable Network Graphic	.png
Images	Tagged Image File	.tif

A very good source for file types and extensions:

<http://fileinfo.com>



Transactional Data vs. Hypermedia for Data in Motion

Unstructured Data Formats and File Types

A few common examples of file types and extensions for unstructured data.

Type	Name	File Extension
Video	Audio Video Interleave File	.avi
Video	Flash Video File	.flv
Video	Apple QuickTime Movie	.mov
Video	MPEG-4 Video File	.mp4
Video	MPEG Video File	.mpg
Video	Flash Movie	.swf
Video	Windows Media Video File	.wmv

A very good source for file types and extensions:

<http://fileinfo.com>



Transactional Data vs. Hypermedia for Data in Motion

Unstructured Data Formats and File Types

A few common examples of file types and extensions for unstructured data.

Type	Name	File Extension
Audio	Advanced Audio Coding File	.aac
Audio	Audio Interchange File Format	.aif
Audio	Interchange File Format	.iff
Audio	MP3 Audio File	.mp3
Audio	MPEG-2 Audio File	.mpa
Audio	Real Audio File	.ra
Audio	Wave Audio File	.wav

A very good source for file types and extensions:

<http://fileinfo.com>



The Challenges of Unstructured Data



Transactional Data vs. Hypermedia for Data in Motion

The Challenges of Unstructured Data

- The difficulties with SOA and defining transactions with unstructured are several.
- Primarily, these challenges include:
 - ❑ The supporting technology (file size, bandwidth, download speed)
 - ❑ Classification for discovery (taxonomies and tagging)
 - ❑ Numerous and varied formats and encodings
 - ❑ Limited metadata and the lack of “structure”



Transactional Data vs. Hypermedia for Data in Motion

The Challenges of Unstructured Data

- So how do we resolve these challenges and what is the role of data ?
- From a SOA perspective, there are a few different approaches to management, exchanging and consuming unstructured data.
- The first is to recognize that not all SOA Service types will apply well to all types of data.
- As we have seen so far, there are significant differences between structured data (elements, types, etc.) and unstructured data (images, video, audio, email, documents, and so on ...)
- When conducting enterprise business that is transactional, the ability to describe and define data, and to initiate and complete business transactions is critical.
- However, when presenting information such as that of a Web page, or a catalog, or to exchange information that is not transactional (Email and documents), the notion of strong typing or completing a business transaction may be less important.



Web Services and ReST



Transactional Data vs. Hypermedia for Data in Motion

Web Services

- Web Services can be thought of a set of servicing behavior that is exposed to consumers, using industry standards of: XML, XML Schemas (XSD), SOAP and WSDL
- Although SOA has evolved to become a recognized approach for implementing business processes and technology solutioning, there is no explicit requirement that SOA uses Web Services.
- However, Web Services work very well for business transactions and the movement, manipulation and exchange of data in motion that is comprised of structured data.
- That is, business transactions, such as: accepting new orders for products, maintaining product data, defining and maintaining customers, accepting payment, etc.
- The data of a Web Services transaction, is carried in the body of a SOAP message, and includes one or more, named and defined elements.
- The metadata for these SOAP Web Services message elements is describe using XML Schemas (XSD)



Transactional Data vs. Hypermedia for Data in Motion

Web Services Standards for SOA

- **XML – eXtensible Markup Language**

- XML is used as the tagging and message encoding scheme.
- XML is generally platform agnostic and can use Unicode
- XML is both extensible and flexible

- **XSD - XML Schemas**

- XSD is a metadata language for XML.
- XSD describes and constrains the structure, names and content of a referencing XML message (or other XML artifact).
- XSD provides declarations and rules to describe things like: data types, structural relationships, cardinality, valid values, abstraction, polymorphism, derivation, and inheritance.



Transactional Data vs. Hypermedia for Data in Motion

Web Services Standards for SOA

- **SOAP – Simple Object Access Protocol**
 - SOAP is often described as an envelope framework
 - SOAP is the over-all message encapsulation and encoding scheme
 - SOAP relies on XML encoding and as a result, is also largely platform agnostic
 - The SOAP “body” contains the message context

```
<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
                    xmlns:SOAP-ENC="http://schemas.xmlsoap.org/soap/encoding/"
                    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
                    xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <SOAP-ENV:Body>
    <m:Request xmlns:m="http://Widget-Example.com/Inventory">
      <m:ItemID>5678</m:ItemID>
      <m:Desc>Blue Stick Pen</m:Desc>
      <m:Version>2</m:Version>
      <m:Length>8.0 Inch</m:Length>
      <m:Color>Blue</m:Color>
      <m:Diameter>0.25 Inch</m:Diameter>
      <m:Weight>1.5 oz</m:Weight>
    </m:Request>
  </SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```



Transactional Data vs. Hypermedia for Data in Motion

Web Services Standards for SOA

- **WSDL – Web Services Description Language**
 - WSDL defines the over-all and technical service interface:
 - ❑ Service Name and Location
 - ❑ The Message format (via XML Schemas)
 - ❑ The protocol and encoding of the message
 - ❑ The Message Exchange Pattern (MEP)
 - ❑ The Operations that are exposed by the service
 - WSDL either intrinsically embeds, or references via “import” - one or more XML Schemas



Transactional Data vs. Hypermedia for Data in Motion

Web Services

- The role of the Data Architect in a SOA Web Service environment, should include some method of defining consistent metadata and defining and exploiting an enterprise canonical model.
 - ❑ Element Name (tag names)
 - ❑ Data Type
 - ❑ Length
 - ❑ Decimal Scale and Precision
 - ❑ Modality
 - ❑ Value Values
 - ❑ And so on ...
- The Data Architect should also participate in the mapping of message elements to the source and target persistent data sources.



Transactional Data vs. Hypermedia for Data in Motion

ReSTful Services

- An alternative to SOAP Web Services are ReSTful services.
- ReST stands for “Representational State Transfer”.
- ReST was defined as an architectural Web and servicing style by Roy Fielding in his PhD dissertation ⁽¹⁾ .
- Dr. Fielding described ReST as being similar to how the World Wide Web works today.
- When you request a Web page from your browser, you are actually sending a request for a “resource” (a complete thing that is identified by a URI and/or located at a URL).
- If the Web page is found at the URL you specific in your browser address line, that resource is then returned to your browser and it is rendered or otherwise consumed.
- This simple analogy works very well and is proven by millions of Web site and Web pages.

1. Roy Thomas Fielding, “Architectural Styles and the Design of Network-based Software Architectures”, University of California, Irvine, Dissertation, 2000,
<http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>



Transactional Data vs. Hypermedia for Data in Motion

ReSTful Services

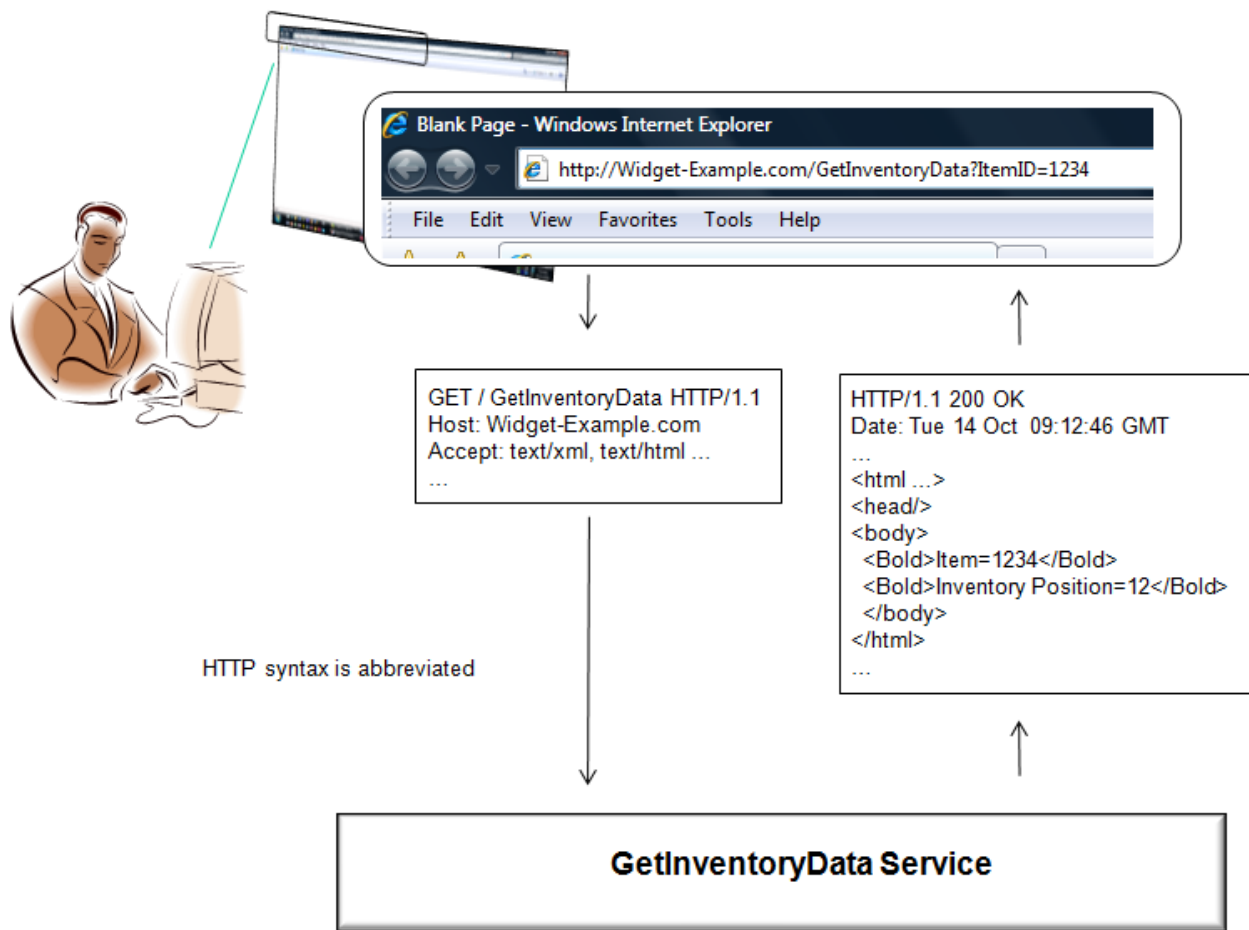


Figure 3.5
SOA and Web Services Interface Design
Morgan Kaufmann © 2009
By James Bean



Transactional Data vs. Hypermedia for Data in Motion

ReSTful Services

- If we think about unstructured data and especially Hypermedia, the ReST model is generally a very good fit.
- Images, Audio Files and Video Files are all “resources”, and can be easily requested from a Web server and file management system, as a unit.
- They can also be downloaded as a “resource” or file using protocols like FTP (File Transfer Protocol)
- These same “resources” (images, etc.) can also be linked and referenced from within a Web page, and resolved as a single resource for rendering by a browser.
- From a business transaction perspective, we could have a Web site that includes Web pages describing our products and services, and also allowing for downloading of resources (think about examples like Amazon.com, Walmart.com, Sears.com and so on).
- However, there are also some challenges ...



Transactional Data vs. Hypermedia for Data in Motion

ReSTful Services

- It is difficult to think of, architect, design, develop and manage a “resource” as a set of individually defined and strongly typed elements.
- While we can present (display, render, download) a resource as a file, it become very difficult to conduct business with this same ReSTful pattern.
- Effectively conducting business with transactions such as Maintain Customer Data, requires very explicit pieces of data and their defining metadata.
- Data elements like:
 - ❑ Customer Name
 - ❑ Customer Address
 - ❑ Customer Telephone
 - ❑ Maximum Credit Amount
 - ❑ Amount Currently Due



Transactional Data vs. Hypermedia for Data in Motion

ReSTful Services

- While ReST is very good for identifying, exchanging and downloading “resources” such as image files, video files, and audio files, most business transactions require more discretely defined data elements.
- These individual data elements are combined together in a SOA to represent a message and with a servicing context of a business function, operation and behavior. We can think of this as a “Uniform Interface”.
- Dr. Fielding even describes in his dissertation that a ReSTful interaction and a resource do not always meet the need for a uniform interface ⁽²⁾
- This is where a solid understanding of the business use case, the type of transaction, and the type of data become critical.

2. Roy Thomas Fielding, “Architectural Styles and the Design of Network-based Software Architectures”, Section 5.1.5, Uniform Interface”, University of California, Irvine, Dissertation, 2000, <http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>



Data in Motion



Transactional Data vs. Hypermedia for Data in Motion

Data in Motion

- If we think of the messages that are exchanged between services and service consumers, we have what is known as “Data in Motion”.
- That is, the data contained within the message is moved from one place to another.
- So ... For Data in Motion, which is better – Web Services or ReST ?
- Not a simple answer. As I hope we have seen so far, both Web Services and SOA can play a critical and effective role in a SOA.
- Where our services will need to represent business transactions, Web Services are most often the better choice.
 - We will have a well-defined set of metadata
 - We will be able to manage the message contents as a transaction, as well as individual and discrete data elements and values
 - We can easily manipulate individual data values of the transaction to conduct business
 - We will have transaction and interface consistency



Transactional Data vs. Hypermedia for Data in Motion

Data in Motion

- Where our services will need to represent , contain or reference Hypermedia or unstructured data as “resources”, then ReSTful Services are most often the better choice.
 - We will have a simple and proven method of exchanging, moving and downloading the data (“resource”)
 - We will have a set of well-defined resource and file types
 - We can port these resources to applications such as browsers and media players for consumption and rendering
- Is it that simple ? That is all we have to think about and manage ?
- Not really. There is more. In some cases – much more !



Metadata and Tagged Content



Transactional Data vs. Hypermedia for Data in Motion

Metadata and Tagged Content

- We have covered SOA, Services and Data in Motion quite well.
- We know that data and metadata play critical roles in a SOA, as do the service interface and the style of servicing for Data in Motion.
- However, our data and our “resources” (hypermedia and unstructured data) originate from somewhere and for many interactions, they will be resolved at a target.
- That is, Data at Rest
- Many Business Enterprise will also have to manage some combination of structured and unstructured data.
- This might be handled at an application level by: a traditional database, a content management system, or a file management system



Transactional Data vs. Hypermedia for Data in Motion

Metadata and Tagged Content

- For Hypermedia and unstructured data, this can be a challenge.
- Services that will expose unstructured data will need some form of identifier and a resolvable location to host it.
- Unstructured data that is persisted will need a method of identification and also a method of description to support discovery.
- This introduces yet another type of metadata – a classification scheme, often described as “tagging”.
- Tagging, allows us to assign a set of descriptors and search terms to our unstructured data.
- Using a simple hypermedia example of a music recording, knowing the file format and name of the recording is critically important. However, if we had a catalog of thousands of recordings, we might also want to be able to search by: Artist, Data Recorded, Copyright Holder, Genre, and Run Time.



Transactional Data vs. Hypermedia for Data in Motion

Metadata and Tagged Content

- The importance here is to consider alternative forms of metadata and how that metadata might fit unstructured data.
- For the business enterprise, we might have to manage other types of metadata such as Email or Documents.
- Here again, we need a method of classification. We might need to know the Author, Data Composed, Data Sent, Title and so on ...
- Sometime, take a look at a Microsoft Word document, and open the properties. Then open the advanced properties and look at the various options. These are predefined and you can define you own as a well.



Transactional Data vs. Hypermedia for Data in Motion

Metadata and Tagged Content

- As with any form of metadata, consistency becomes important.
- That classifications we might find important to describe Music Audio files could be very different from the classifications we use to describe Email.
- Even within one media type, we might have some classifications that are mandatory and others that are optional.
- As the classification are valued, we also need consistency. Using the earlier example of a music recording, we probably would not want 2 separate values of a Genre Classifier for “Country” such as “Country” and “CNTRY”.
- At the same time, if we need to consider that one of the use cases includes human discovery, we might want to allow multiple classifications of a general type for a single instance of our unstructured data. As an example, we might want to classify a single music recording with 2 or 3 Genres as: “Rock”, “Country” and “Rock-a-Billy”.



Transactional Data vs. Hypermedia for Data in Motion

Metadata and Tagged Content

- To apply classification metadata consistently, we need to embrace the notion of controlled vocabulary and taxonomy.
- The controlled vocabulary defines sets of taxonomies and allows for things like differences in language. It also defines the set of concepts under which our taxonomy will be developed.
- The taxonomy is more of a hierarchy of nouns, verbs, adjectives and adverbs that also include synonyms.

ANSI/NISO Z39.19 - Guidelines for the Construction, Format,
and Management of Monolingual Controlled Vocabularies 2005
(R2010)

http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key=7cc9b583cb5a62e8c15d3099e0bb46bbae9cf38a

Linnaean Taxonomy, Wikipedia

http://en.wikipedia.org/wiki/Linnaean_taxonomy



The Role of the Data Architect



Transactional Data vs. Hypermedia for Data in Motion

The Role of the Data Architect

- The Data Architect should be deeply involved in the definition and maintenance of all metadata – both data at rest and data in motion.
- However, Data in Motion is NOT the same as Data at Rest. There are many similarities and several synergies. Metadata being the most obvious.
- There are some deep skills required to fully understand SOA, Services, XML Schemas, and the Interface Contract.
- Attempting to apply traditional metadata practices for persistent data to a SOA and Servicing paradigm can result in a very ineffective implementation.
- Each enterprise has a different organizational structure and your role might vary depending upon your approach to SOA and Services.



Transactional Data vs. Hypermedia for Data in Motion

The Role of the Data Architect

- First and foremost – Education !
- I recommend that an minimum, the Data Architect gains tactical, hands-on experience with:
 - ❑ Canonical Models and Metadata
 - ❑ SOA Concepts, Principles and Patterns
 - ❑ Service Message Exchange Patterns (MEP)
 - ❑ XML
 - ❑ XML Schemas (XSD)
 - ❑ Object Oriented Architecture and UML
- As additional knowledge, I recommend that the Data Architect also gains knowledge of:
 - ❑ Controlled Vocabulary, Taxonomy
 - ❑ SOAP
 - ❑ WSDL
 - ❑ ESB
 - ❑ BPM
 - ❑ Service Registry



Transactional Data vs. Hypermedia for Data in Motion

The Role of the Data Architect

- Once the appropriate skills have been gained, the Data Architect should look to engage in two critical areas for SOA and for Unstructured Data:
 1. To either define or participate in the definition of the Service Interface
 2. To either define or participate in the definition of the Tagging Scheme for Unstructured Data



THANK YOU !

Questions ?



References and Bibliography



Transactional Data vs. Hypermedia for Data in Motion

References and Bibliography

Images as clipart originated from Microsoft and the “Clip Art on Office Online” site
<http://office.microsoft.com/en-us/images/>

File Types and Extensions
<http://fileinfo.com>

SOA and Web Services Interface Design
Morgan Kaufmann © 2009
By James Bean

Roy Thomas Fielding, “Architectural Styles and the Design of Network-based Software Architectures”, University of California, Irvine, Dissertation, 2000,
<http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>

ANSI/NISO Z39.19 - Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies
2005 (R2010)
http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key=7cc9b583cb5a62e8c15d3099e0bb46bbae9cf38a

Linnaean Taxonomy, Wikipedia
http://en.wikipedia.org/wiki/Linnaean_taxonomy

